5-2023

piRNA expression in regenerative tissue of Octopus bimaculoides

Bailey Ervin

piRNA expression in regenerative tissue of *Octopus bimaculoides*

by

Bailey Ervin

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

May 2023

Approved by:

Alex Flynt, Ph.D., Thesis Advisor,
School of Biological, Environmental and Earth Sciences

Jake Schaefer, Ph.D., Director,
School of Biological, Environmental and Earth Sciences

Sabine Heinhorst, Ph.D., Dean
Honors College

ABSTRACT

Tissue regeneration is present in varying capacities across the animal kingdom. Animals such as *Hydra* and planarians have the capacity to regenerate entire bodies from extremely small sections of amputated tissue. Others, such as humans, have restricted capacities of regeneration, especially in terms of full appendages and specialized tissues such as cardiac and nervous tissue. One of the primary goals of studying regeneration in other organisms is to achieve the development of regenerative medicine. Interaction of P-element induced WImpy testis (PIWI) proteins and PIWI-interacting RNAs (piRNAs) have been implicated in germline genome maintenance, as well as transposable element silencing. Research has also connected PIWI protein expression to regeneration in model organisms. *Octopus bimaculoides* is a marine animal of Phylum Mollusca that exhibits great regenerative abilities. In this research study, piRNA expression was examined in the regenerated tentacles of *O. bimaculoides*, and its somatic tissue used as the control. RepeatMasker analysis showed that piRNAs were targeting repeated elements, most notably DNA transposons, long tandem repeats (LTRs), and long interspersed nuclear elements (LINEs). Gene ontology analysis showed that piRNAs were targeting genes implicated in the regulation of transcription, cell communication, signal transduction, and intrinsic and integral components of the cell membrane.

Keywords: piRNA, PIWI, transposons, *Octopus bimaculoides*

DEDICATION

I would like to dedicate this thesis to my parents, Granny, siblings, and extended family for always supporting me as a strive to further my education as an undergraduate and in my pursuit of medical school, and to my best friends Griffin, David, Brennan, Ashton, and Jacob for their support, being there for me, and for reminding me to have fun along the way.

ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Dr. Alex Flynt, for providing me the opportunity to conduct such meaningful research in his lab. His expertise and insight were truly instrumental to the success of this thesis. I would also like to thank the post-baccalaureate student that I worked with throughout this research, Beatriz Zancanela. I greatly appreciate Beatriz's selflessness, kindness, and prowess which was shown through the many hours that she spent helping me for the duration of this project. I must also thank the University of Southern Mississippi, the School of Biological, Environmental, and Earth Sciences, and the Honors College for providing me with the opportunity to participate in research of this caliber.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

Ago	Argonaute
Ago3	Argonaute 3
Aub	Aubergine
BP	biological process
C	carboxyl terminus
CC	cellular component
cDNA	complementary DNA
DAG	directed acyclic graph
DGE	differential gene expression
DNA	Deoxyribonucleic Acid
dSetDB1	*Drosophilia* SetDB1 protein
GO	gene ontology
H	histone
H3K9	methylate lysine 9 residue in histone 3
H3K9me3	histone 3 lysine 9 trimethylation
K	lysine
kD	kilodalton
LINEs	long interspersed nuclear elements
LTR	long terminal repeats
Me	methylation
MF	molecular factor
miRNA	micro RNA

mRNA	messenger RNA
N	Nitrogen (Amino terminal)
Nt	nucleotides
Nxf2	Nuclear export factor 2
Nxt1	Nuclear transport factor 2 – like export factor 1
PAZ	PIWI-Argonaute-Zwille
pH	potential of hydrogen
piRNA	PIWI-interacting RNA
PIWI	P-element induced WImpy testis
Piwil1	Piwi-like RNA-mediated gene silencing 1
Piwil2	Piwi-like RNA-mediated gene silencing 2
Ppt	parts per thousand
rasiRNAs	repeat-associated small interfering RNAs
RNA	Ribonucleic acid
RNAi	RNA interference
RNA Pol II	RNA Polymerase II
RNA-seq	RNA sequencing
RPKM	read per kilobase per million mapped reads
RSEM	expectation maximization
SINEs	short interspersed nuclear elements
siRNA	small interfering RNA
snRNAs	small noncoding RNAs
Zuc	Zucchi

CHAPTER I: INTRODUCTION

Regeneration can be observed, even if only in rudimentary forms, in all animals. This suggests that regenerative abilities may have evolved ancestrally in metazoans. All animals require some level of regeneration or self renewal to replace dead, dying, or mutated cells (Poss, 2010). However, there is immense variability in regenerative capabilities across the many different taxa within the animal kingdom. Members of the genus *Hydra* may perhaps exhibit the most extensive regenerative capabilities of all animals (Poss, 2010). They have been known to regenerate entire individuals from just a small piece of tissue or even a small aggregate of cells (Poss, 2010). Animals such as humans and most mammals lie on the opposite side of the spectrum of regenerative capacity. The regenerative abilities of these animals vary depending on factors such as location and tissue type. Tissues such as blood and skin have relatively high regenerative abilities due to their respective stem cell populations. Other tissues in these animals such as cardiac tissue, nervous tissue, and major appendages have alarmingly little regenerative capabilities (Poss, 2010).

Regeneration has been extensively studied in model animals, such as planarians. Planarians exhibit extraordinary levels of regenerative capacity. Similar to *Hydra*, planarians can regenerate entire bodies from just small pieces of tissue (Poss, 2010). The mechanisms of planarian regeneration are morphallaxis and epimorphosis (Handberg-Thorsager et al., 2008; Pellettieri, 2019). Morphallaxis is regeneration through remodeling of existing tissue, whereas epimorphosis is regeneration through blastema formation (Pelletieri, 2019). The blastema is a regenerative bulb of cells called neoblasts, which are totipotent stem cells (Rossi et al., 2008). The exact mechanisms by which

blastemas contribute to the regenerative process are not yet fully understood. PIWI proteins are members of the Argonaute protein family that have been implicated in the process of regeneration (Carmell et al., 2002). PIWI proteins have also been found to interact with PIWI-interacting RNAs (piRNAs) in complexes that function in the silencing of transposable elements and maintenance of the germline (Yamaguchi et al., 2020). It has been established more recently that PIWI-piRNA complexes likely serve functions in regenerative tissue. Many mollusks of the Class Cephalopoda, such as the octopus species *Octopus vulgaris and Octopus bimaculoides*, exhibit impressive regenerative abilities. They are especially known for their ability to regenerate amputated appendages, or tentacles (Shaw et al., 2016).

This research revolving around regeneration in model organisms is laying the groundwork for regenerative medicine. The aim of regenerative medicine is to use regeneration to heal ailments that are otherwise extremely difficult to treat, such as regeneration of a limb lost in combat or a diseased organ that is failing to fulfill its normal function. The opportunities in regenerative medicine are near boundless. However, scientists must first develop a comprehensive understanding of regeneration in other organisms. This will hopefully foster breakthroughs in regenerative medicine and lead to its application to higher vertebrates in the near future. The aim of this study was to show the levels of differential expression of piRNAs between somatic tissue and regenerative tissue of *O. bimaculoides*, working towards establishing the role of piRNAs in regenerative tissue.

Hypothesis

Differential expression was expected to be seen between the regenerative and control tissues, as the there should be higher expression of genes related to growth and cell cycle control in regenerative tissue.

CHAPTER II: LITERATURE REVIEW

Tissue Regeneration

Regeneration refers to the replacement of tissue lost as a result of injury (Poss, 2010). There are two primary types of regeneration: homeostatic and facultative (Poss, 2010). The latter refers to regeneration resulting from considerable injury, whereas the former refers to regeneration of cells lost during normal biologic function (Poss, 2010). One of the primary purposes of conducting research on regeneration is to contribute to the field of regenerative medicine. This can, for example, be accomplished by acquiring knowledge regarding the stimulation and manipulation of stem cells.

Planarians are one of the most understood, heavily researched model organisms of regeneration in invertebrates. Planarians belong to Class Turbellaria of Phylum Platyhelminthes (Saló & Baguñà, 2002). They are dorsoventrally flattened, have three germ layers, and lack a coelom, segments, skeleton, and structures responsible for respiration and circulation (Saló & Baguñà, 2002). The mechanisms responsible for their regenerative abilities have been suggested to be morphallaxis and epimorphosis (Handberg-Thorsager et al., 2008; Pellettieri, 2019). The blastema is derived from stem cells formally known as neoblasts, which are totipotent adult stem cells located in the parenchyma that have the capacity to multiply in asexual strains (Rossi et al., 2008). Within 30 minutes of amputation, the wound site is covered by an epithelial layer (Pellettieri, 2019). The cells that will form the regenerated tissue are then formed by a spike in neoblast mitosis in the tissue next to the blastema (Pellettieri, 2019). These cells cease mitosis upon arrival to the blastema area (Pellettieri, 2019). Grafting experiments

have shown that determination occurs rapidly following wound healing (first 24 hrs in the head, 36 hrs in the pharynx) (Pellettieri, 2019).

Zebrafish (*Danio rerio*) are one of the most understood, heavily researched model organisms of regeneration in vertebrates. *Danio rerio* are members of Phylum Chordata and Class Teleostei. They have been known to regenerate fins, scales, heart tissue, and more (Poss et al., 2003). It has been determined that zebrafish utilize epimorphosis to regenerate amputated fins. Within one to three hours of caudal fin amputation, the wound site is covered by an epithelial layer (Poss et al., 2003). The cells of this layer do not undergo proliferation, but rather migration (Poss et al., 2003). An early molecular marker for fin regeneration is the presence of $_{\beta}$-catenin in the wound epithelial cells (Poss et al., 2003). The blastema is formed following closely after wound closing (Poss et al., 2003). As in planarians, the blastema is derived from pluripotent stem cells (Poss et al., 2003). Between two and four days following amputation, the blastema transitions from formation to a regenerative outgrowth (Poss et al., 2003). With this transition, the rate of blastemal cell cycles drastically increases (Poss et al., 2003).

Octopus bimaculoides

Octopus bimaculoides was first described by Grace E. Pickford and Bayard H. McConnaughey in 1949 (Pickford & McConnaughey, 1949). It is a member of Phylum Mollusca and Class Cephalopoda (Allock et al., 2018). This animal is also referred to by its common names, the California two-spot octopus or bimac octopus. All three of its names refer to its exhibition of two false eyes, formally known as ocelli, below each of its eyes (Hofmeister, 2015). *Octopus bimaculoides* is native to the coasts of California and Mexico (Jereb & Roper, 2010). They usually reside on rocky reefs or sand in intertidal or

shallow subtidal zones (Allock et al., 2018). The average size of its mantle is about 17.5 cm, and the average length of its tentacles is about 58 cm. A hallmark of this animal is its highly evolved ability of crypsis, or camouflage. One of the common color combinations under normal conditions is grey with yellow spots. The life expectancy of *O. bimaculoides* is roughly two years. Its genome was sequenced in 2015 (Albertin et al., 2015).

Regeneration in cephalopods has been studied for over a century. Although regeneration in *O. bimaculoides* specifically has not been extensively studied, regeneration in *Octopus vulgaris*, a close relative of *O. bimaculoides*, has been (Shaw et al., 2016). *Octopus vulgaris*, also known as the common octopus, has been heavily examined historically as a model cephalopod. In 1920, Mathilde M. Lange studied regeneration in *O. vulgaris* (Lange, 1920). She determined distinct morphological changes, such as the use of an epithelial lining in closure of the wound and blastema formation, throughout the regeneration of amputated tentacles (Lange, 1920). This field of research has evolved since Lange's studies through the acquisition of new information and technology. Using high-resolution ultrasound imaging, Shaw et al. found that epithelium was contracting to cover the wound within 24 hrs of amputation (Shaw et al., 2016). Their findings also included the use of a "cell plug" and apoptosis in early wound closure and healing (Shaw et al., 2016). Numerous characteristics of regeneration in cephalopods have yet to be studied, such as the molecular and cellular machinery involved (Imperadore & Fiorito, 2018).

PIWI Proteins

The Argonaute (Ago) proteins constitute a conserved family of proteins that have been linked to RNA interference (RNAi) and similar processes in numerous organisms (Carmell et al., 2002). They have also been found to play critical roles in development and stem cell determination (Carmell et al., 2002). They weigh approximately 100 kilodaltons (kD) and are profoundly basic (Carmell et al., 2002). All Argonaute proteins contain two common domains: PAZ and PIWI (Cerutti et al., 2000). The PAZ domain is located toward the N terminus. It is composed of 130 amino acids and theorized to function in interprotein interaction (Cerutti et al., 2000; Bernstein et al., 2001). The PIWI domain is located toward the C terminus. It is composed of 300 amino acids and its specific function has not been successfully established. The Argonaute family of proteins can be further divided into two groups: Argonaute proteins and PIWI proteins (Carmell et al., 2002). The Argonaute proteins have been implicated in gene silencing through interaction with micro RNAs (miRNAs) and small interfering (siRNAs). PIWI proteins have been found to play a critical role in the maintenance of germline integrity.

PIWI expression in somatic tissue

The somatic stem cells of organisms with high regenerative capabilities, such as planarians, contain high levels of PIWI proteins (Reddien et al., 2005). The regenerative capabilities of these organisms are lost upon loss of PIWI proteins from the somatic stem cells (Reddien et al., 2005). Studies have demonstrated the crucial role of PIWI proteins in the somatic stem cells of planarians. The totipotent stem cells of planarians, known as neoblasts, have the ability to repopulate both germline and somatic lineages (Ross et al., 2014). PIWI proteins play a crucial role in the neoblast lineage. For instance, Reddien et

al. (2005) examined the phenotypic implications of the knockdown of PIWI-encoding messenger RNAs (mRNAs). They found that this knockdown of PIWI-encoding mRNAs eliminated the planarian's capacity to regenerate body parts and caused the loss of their tissue maintenance abilities, ultimately causing death (Reddien et al., 2005). *Drosophilia* express the PIWI protein in all ovarian somatic cells and in early somatic cells of the testis (Cox et al., 2000). Brower-Toland et al. found that the polytene chromosomes of the salivary gland bound PIWI (Brower-Toland et al., 2007). It has also been established that PIWI proteins function in the head region of *Drosophilia* (Ross et al., 2014).

PIWI expression in germline tissue

PIWI expression in germline tissue has been extensively studied. PIWI expression in germline tissue is critical to germline determination, integrity, maintenance, and gametogenesis (Thomson & Lin, 2009). Thomson & Lin found PIWI proteins associated with polar granules in *Drosophilia* (Thomson & Lin, 2009). Polar granules are electron dense structures heavily involved in the production of gametes (Thomson & Lin, 2009). PIWI proteins are widely known for their interactions with piRNAs. These interactions are crucial to the process of silencing transposable elements (Brennecke et al., 2007). Loss of PIWI protein function has drastic consequences. PIWI protein gene mutations in *Drosophilia* have been implicated in sterility and loss of germline stem cells (Cox et al., 1998). Additionally, loss of PIWI protein function would cause a drastic increase in the accumulation of transposable element transcripts due to the lost transposable element silencing mechanism.

PIWI expression in Octopus bimaculoides

PIWI proteins and their interaction with piRNAs are critical to regulation of transposons on both post-transcriptional and transcriptional levels. There are only a handful of functional descriptions of PIWI proteins and piRNAs in the Phylum Molluska (Jehn et al., 2018). Most of these descriptions stem from studies conducted on *Aplysia californica* (sea slug), *Chlamys farreri* (Farrer's scallop), and *Nucella lapillus* (dog whelk) (Jehn et al., 2018). In 2018, Waldron et al. established that piRNA-like small RNAs matching transposon sequences are expressed in somatic tissue of *Nucella lapillus* (Waldron et al., 2018). Unfortunately, it is not possible to determine whether this is a conserved or lineage-based feature of PIWI-piRNA interaction in mollusks. Using PIWI protein sequence data from *Biomphalaria glabrata, Aplysia californica, Crassostrea gigas, Crassostrea virginica, Mizuhopectin yessoensis,* and *Octopus bimaculoides*, Jehn et al. found that PIWI proteins Piwil1 and Piwil2 are conserved in mollusks and are orthologous to Piwil1 and Piwil2 in vertebrates (Jehn et al., 2018).

piRNAs

PIWI-interacting RNAs (piRNAs) are small noncoding RNAs (snRNAs) that are usually between 26 and 31 nucleotides (nt) long. piRNAs may be derived from either repeated or complex DNA segments (Aravin et al., 2007). A subset of piRNAs derived from transposons and other repeat elements are known as repeat associated small interfering RNAs (rasiRNAs) (Klattenhoff & Theurkauf, 2008). Research conducted on mice, *Drosophilia*, and zebrafish has established the critical role of piRNAs in germline development (Klattenhoff & Theurkauf, 2008). Some proteins linked to piRNA expression have been found to play a role in somatic gene expression control, and even

learning (Klattenhoff & Theurkauf, 2008). The results of studies surrounding the topic of piRNAs have established a broad range of biological processes they may affect.

piRNAs are usually derived from regions of the genome termed piRNA clusters that are dense in transposable element sequences (Huang et al., 2017). In *Drosophilia melanogaster*, piRNA clusters are subdivided into two classes: uni-strand and dual-strand (Huang et al., 2017). Uni-strand piRNA transcription utilizes promoters and produce 5' capped/3' polyadenylated piRNAs, while dual-strand piRNA transcription does not use promoters and does not produce 3' polyadenylated piRNAs. Some individual transposons and other repeated elements, such as the untranslated regions of protein-coding genes, have also been shown to produce piRNAs (Huang et al., 2017). The histone 3 lysine 9 tri-methylation (H3K9me3) mark is necessary for piRNA expression (Huang et al., 2017).

piRNA biogenesis is the process by which piRNAs are created (Fig. 1). There are two types of piRNA biogenesis: primary piRNA biogenesis and the Ping-Pong cycle. Primary piRNA biogenesis generates primary piRNAs through a process partially outlined above. The first step in this type of genesis is transcription of piRNA sequences to create the precursors of piRNA (Fig. 1)(Huang et al., 2017). Once transcribed, these precursors migrate to the cytoplasm for processing by an enzyme on the mitochondrial membrane called Zucchini (Zuc)(Fig. 1). Zucchini is an endonuclease that forms the 5' end of a given piRNA by cleaving the precursor (Huang et al., 2017). Following 5' formation, the precursor is transported to a PIWI/Aubergine (Aub) protein complex for 3' trimming (Fig. 1). At this point, the piRNA molecule is considered a mature primary piRNA (Huang et al., 2017).

The next phase of piRNA biogenesis is the Ping-Pong cycle. The Ping-Pong cycle uses primary piRNAs to generate secondary piRNAs (Czech & Hannon, 2016). In the Ping-Pong cycle, transposable element sequences are cleaved by antisense Aubergine-bound piRNAs, resulting in a new sense piRNA script (Fig. 1)(Czech & Hannon, 2016). This sense piRNA script binds to Argonaute3 (Ago3) protein. This sense piRNA script/Ago3 complex targets piRNA cluster transcripts, in turn generating antisense piRNA scripts (Fig. 1)(Czech & Hannon, 2016). These antisense transcripts bind to Aub, and the same cycle repeats.

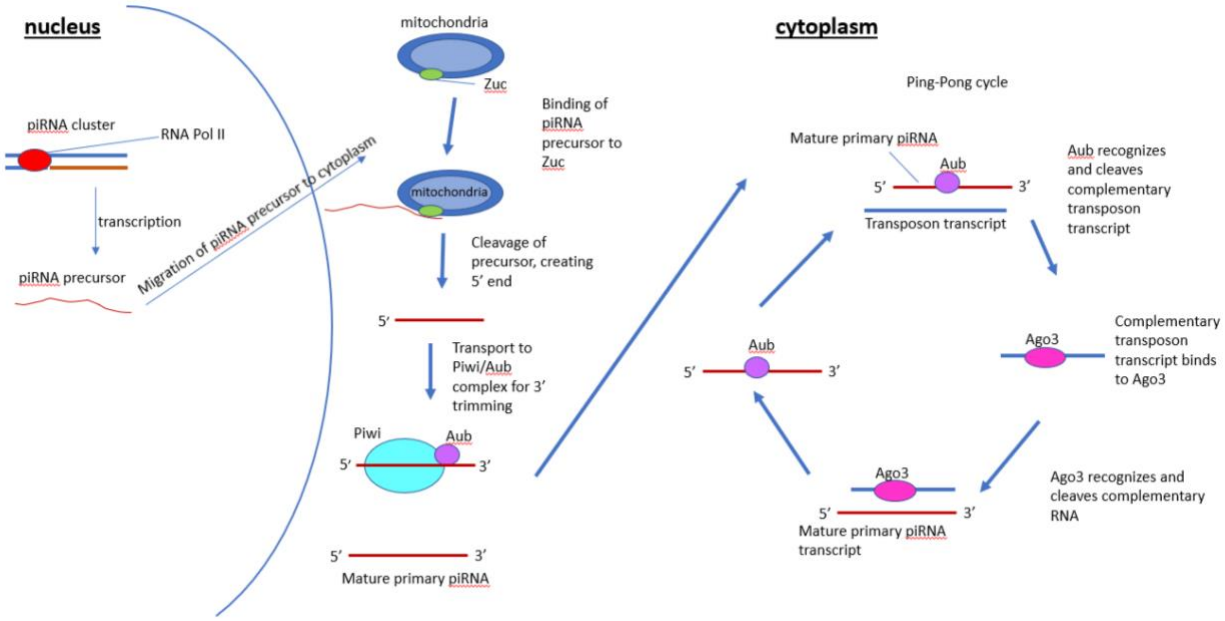

Figure 1: Mechanisms of piRNA biogenesis and the Ping-Pong cycle. This diagram portrays the steps of both primary and secondary piRNA generation. RNA polymerase II (RNA Pol II) transcribes a piRNA precursor from DNA sequence. piRNA precursor migrates to the cytoplasm and interacts with Zucchini (Zuc), an endonuclease in the mitochondrial membrane. Zuc generates the 5' end of the piRNA by cleaving the piRNA precursor. piRNA precursor then undergoes 3' trimming in a PIWI/Aub protein complex. Aub binds to an antisense piRNA script and cleaves a complementary transposon transcript, producing a complementary sense piRNA. Ago3 protein binds the new sense piRNA. Ago3/sense piRNA complex cleaves a complementary piRNA cluster transcript. Result is an antisense piRNA transcript that is bound by Aub, and the process starts again (figure by author).

Transposable elements

Between 33% and 50% of the mammalian genome is derived from transposable elements (Platt et al., 2018). Transposable elements are sequences of DNA that can replicate and insert into new sites in the genome (Percharde et al., 2020). This replication causes accumulation of copies of the elements and expansion of the genome. Transposable elements may be classified as either retrotransposons or DNA transposons (Percharde et al., 2020). Retrotransposons replicate using an RNA intermediate, whereas DNA transposons replicate without the use of an RNA intermediate (Percharde et al., 2020). There are two subgroups within the retrotransposon class: long terminal repeat (LTR) elements and non-long terminal repeat (non-LTR) elements (Percharde et al., 2020). The two subgroups differ in that LTR elements possess LTRs 100-300 base pairs long at their 5' and 3' termini, whereas these structures are not present in non-LTR elements (Percharde et al., 2020). The primary LTR elements are known as endogenous retroviruses, which likely derived from ancestral retroviral infections. The primary non-LTR elements are long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). LINEs are theorized to have derived from bacterial group II introns (Malik et al., 1999). It has been established SINEs are derived from cellular RNAs (Malik et al., 1999).

Transposable elements may contribute to mutational meltdown of populations. Mutational meltdown is a process in which deleterious mutations accumulate in a given population, causing reduced population size and decreased fitness (Platt et al., 2018). This positive feedback loop enhances the accumulation of additional deleterious mutations through genetic drift, eventually causing extinction (Platt et al., 2018), where

lethal transposable element insertions in the germline are removed. On the other hand, equivalent insertions in somatic stem cells are tolerated more often and have been implicated in over 100 diseases, including cancers and neuropathies (Platt et al., 2018). Transposable elements maintain the potential to cause cancer by disrupting protooncogenes or tumor suppressor genes (Platt et al., 2018). Long Interspersed Nuclear Element 1 and Arthrobacter luteus are active retrotransposons that have been implicated in tumorigenesis (Bhat et al., 2022).

Transposable elements have also been found to play a role in gene regulation. They accomplish gene regulation through use of two primary mechanisms: cis-acting control elements and transcriptome modulation through epigenetic control (Bhat et al., 2022). Silencing of transposable elements is one of the chief functions of PIWI proteins and piRNAs. The Aub and Ago3 PIWI proteins of *Drosophilia* function in post-transcriptional silencing (Wang & Lin, 2021). PIWI protein – piRNA complexes in *Drosophilia* also silence transcription by binding incipient transposon transcripts and forcing them into the repressive chromatin state (Wang 7 Lin, 2021). This is accomplished by interactions between the PIWI-piRNA complex and the mediator proteins Asterix and Panoramix (Wang & Lin, 2021). Interaction of PIWI-piRNA complexes with Panoramix activates Nuclear Export Factor 2 (Nxf2) and Nuclear Transport Factor 2 Like Export Factor 1 (Nxt1). These activated factors join dSetDB1 (also known as Eggless) and methylate Lysine 9 residue in histone 3 (H3K9) (Wang & Lin, 2021).

RNA sequencing

RNA sequencing, commonly abbreviated RNA-seq, was developed in the mid-2000s. It is a laboratory technique that employs next-generation sequencing to study the RNA transcripts present in a given sample. RNA-seq examines the transcriptome, which is the complete set of RNA transcripts of a given cell or cell population, to analyze gene expression (Han et al., 2015). RNA-seq is used in techniques such as single nucleotide polymorphism identification and differential gene expression (DGE) analysis. DGE analysis examines how transcript levels differ across samples (Han et al., 2015). DGE analysis utilize methods such as read per kilobase per million mapped reads (RPKM) to quantify transcript abundance (Han et al., 2015). RPKM uses overall mapped read number and gene length to analyze differential expression and normalized read counts (Han et al., 2015). However, RPKM has been criticized because it handles all reads equally without consideration of isoforms. Newer software such as RNA-seq by Expectation Maximization (RSEM) are more accurate in terms of gene and isoform expression levels (Han et al., 2015).

The first step in RNA-seq is extraction of RNA from sample tissue. The target RNAs, in this case snRNAs, of the sample are then enriched, whereas other RNAs present in the sample are reduced (Stark et al., 2019). The target snRNA transcripts are then converted into complementary DNA (cDNA) using an enzyme called reverse transcriptase (Stark et al., 2019). High-throughput platforms such as Illumina are used to create a library of adaptor-ligated sequences (Stark et al., 2019). Illumina is the most heavily used high-throughput platform in research using RNA-seq. The Illumina Nextseq 2000 platform is extremely versatile, as it can be used in RNA-seq to examine spatial

transcriptomics, shotgun metagenomics, single-cell gene expression, and more (Stark et al., 2019). The use of high-throughput platforms yields tens of millions of reads per sample. The final steps of RNA-seq are data analysis. The resultant sequences are aligned against the transcriptome and mapped read/transcriptome overlaps are quantified (Stark et al., 2019). Significant gene expression differences between samples are then determined using statistical analysis. The RNAseq for this research was done in the Molecular Genomics Core Facility at University of Mississippi Medical Center on an Illumina NextSeq 500 instrument. Libraries were created using the Illumina small RNA TruSeq kit.

CHAPTER III: MATERIALS AND METHODS

Upkeep of *Octopus bimaculoides*

The study organism in this experiment was *Octopus bimaculoides*. *Octopus bimaculoides*' natural habitat is the Pacific coast of California. It was necessary that the living conditions of *O. bimaculoides* in the laboratory were matched as closely as possible to those of its natural habitat. The specimen was kept in a 70 gal tank equipped with biofilters. A water pump was attached to a chiller to maintain a tank temperature of 18° C. The specimen was fed one thawed raw shrimp daily. Items such as clay pots and numerous rocks of varying sizes were placed into the tank for shelter and entertainment. The salinity of the tank water was maintained at 35 ppt, and the pH was maintained at 8. Two times every week, 30% of the volume of tank water was changed. Nitrate, nitrogen dioxide, copper, and ammonia levels were kept as close to 0 as possible.

Amputations

The University of Southern Mississippi does not require an IACUC for invertebrates. To conduct the amputations of *O. bimacloides*' tentacles, it was moved from the tank where it was living to a separate tank and submerged underwater. Ethanol was added in 0.25% increments of the total volume of the tank every two minutes until a concentration of 1.5% ethanol was achieved. The specimen was under complete anesthesia after about 14 minutes under these conditions. The amputations were conducted using sharp razor blades. Three tentacles were amputated and allowed to regenerate replicas. Amputated tissue was placed in TRIzol. The same protocol as above was conducted to recover the blastema of regenerated tissue three days post-amputation.

Recovery of the specimen from anesthesia was done by placing it back into its regular tank. Recovery from anesthesia took place in a matter of seconds.

Figure 2: Transverse amputation section of O. bimaculoides *tentacles. The distal ends of three tentacles were cut using a razor blade. Tissue samples were taken by cutting each tentacle against the tank wall (photo from stock images, excision lines by author).*

RNA extractions

To extract RNA from the samples, the sample tissue was first homogenized using a pestle-style tissue grinder. The rest of the extractions were conducted using TRIzol LS and following the manufacturer's instructions. In addition to the manufacturer's instructions, 300 µL of nuclease-free water was added to balance salt present in the tissue samples. The RNA samples were then sent to be sequences using Illumina Nextseq 2000.

Computational data analysis

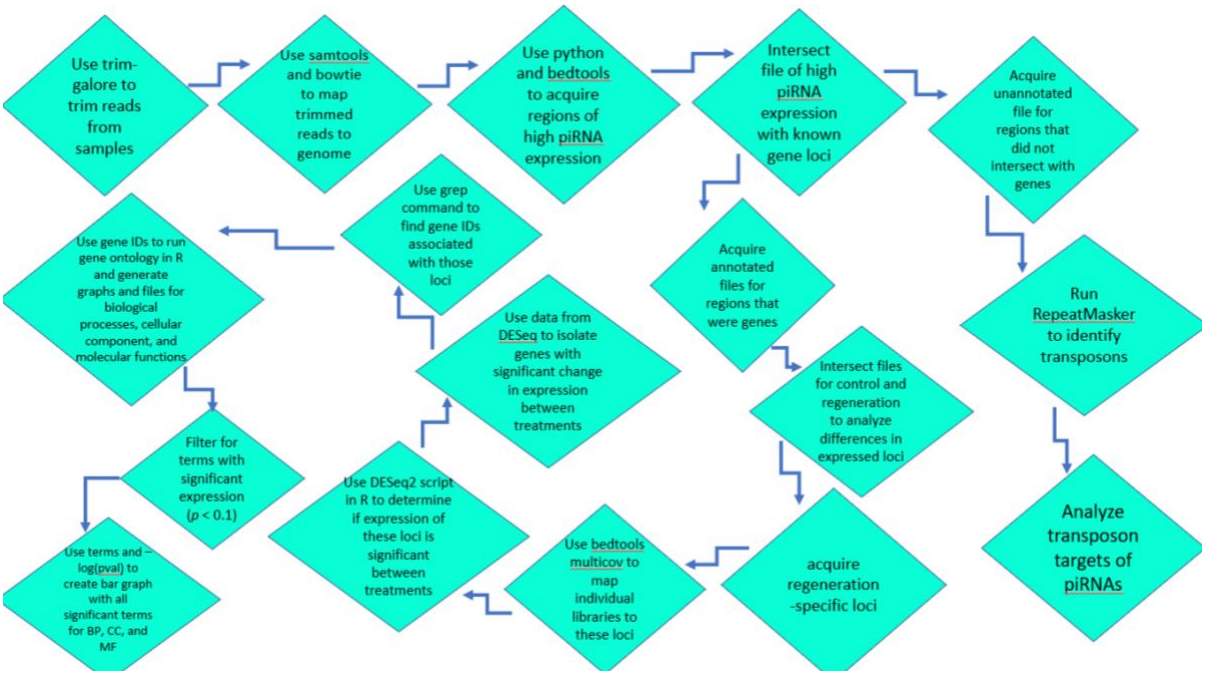

Figure 3: Flowchart of bioinformatics for computational analysis of piRNA expression in somatic and regenerative tissue of Octopus bimaculoides. This flowchart portrays the steps taken to use the RNA sequence received from the O. bimaculoides control and regeneration samples and the O. bimaculoides genome to analyze differential expression of piRNA genes between the two tissue types (figure by author).

Gene ontology analysis

The Gene Ontology (GO) was created with the advent of high-throughput RNA-seq platforms such as Illumina. These RNA-seq technologies produce extremely vast data sets. However, it was difficult to organize and manage the data sets in a way that could be readily used. The GO Consortium was established to define standard ontologies and establish annotations for the GO (Harris et al., 2004). The GO Consortium is composed

of several large databases working in tandem. The GO is a structured vocabulary of terms divided into three non-overlapping ontologies: molecular function (MF), biological process (BP), and cellular component (CC) (Harris et al., 2004). These ontologies portray gene functionality and the relation between terms. The relations are "is a," "part of," "has part," or "regulates" relationships (du Plessis et al., 2011). The "regulates" relationships are further divided in "positively regulates" and "negatively regulates" (du Plessis et al., 2011). The "is a" relationship is used to connect subtypes to their more general counterparts (du Plessis et al., 2011). These relationships establish the borders of Directed Acyclic Graphs (DAG) in which the terms are the nodes (du Plessis et al., 2011). The DAG is the final product of GO analysis. DAGs resemble flowcharts in which the first box is a general domain process. The derived boxes following the general domain process become increasingly specific terms in regard to the function of the given sequence. DAGs present the functions of genes or gene products with increasing specificity and significance using probability values (*p*).

CHAPTER IV: RESULTS

piRNA gene targets in regenerative vs. control tissue

Intersection of files containing sequence of high piRNA expression with known gene loci provided 3382 annotated loci of high piRNA expression in the control (somatic) tissue, and 2153 annotated loci of high piRNA expression in the regenerative tissue (Fig. 4, left). An overlap of 169 loci between the two tissue subsets was found, providing 2071 regeneration-specific loci and 3213 control-specific loci (Fig. 4, left). There were 2768 unannotated loci implicated in regeneration, discussed in depth below under "RepeatMasker results". The total number of loci after filtration for unique loci terms was 1969. DESeq2 was used in R Studio to determine differential expression of regeneration-specific genes. This provided 66 upregulated genes and 45 downregulated genes, depicted in the volcano plot on the right below (Fig. 4, right).

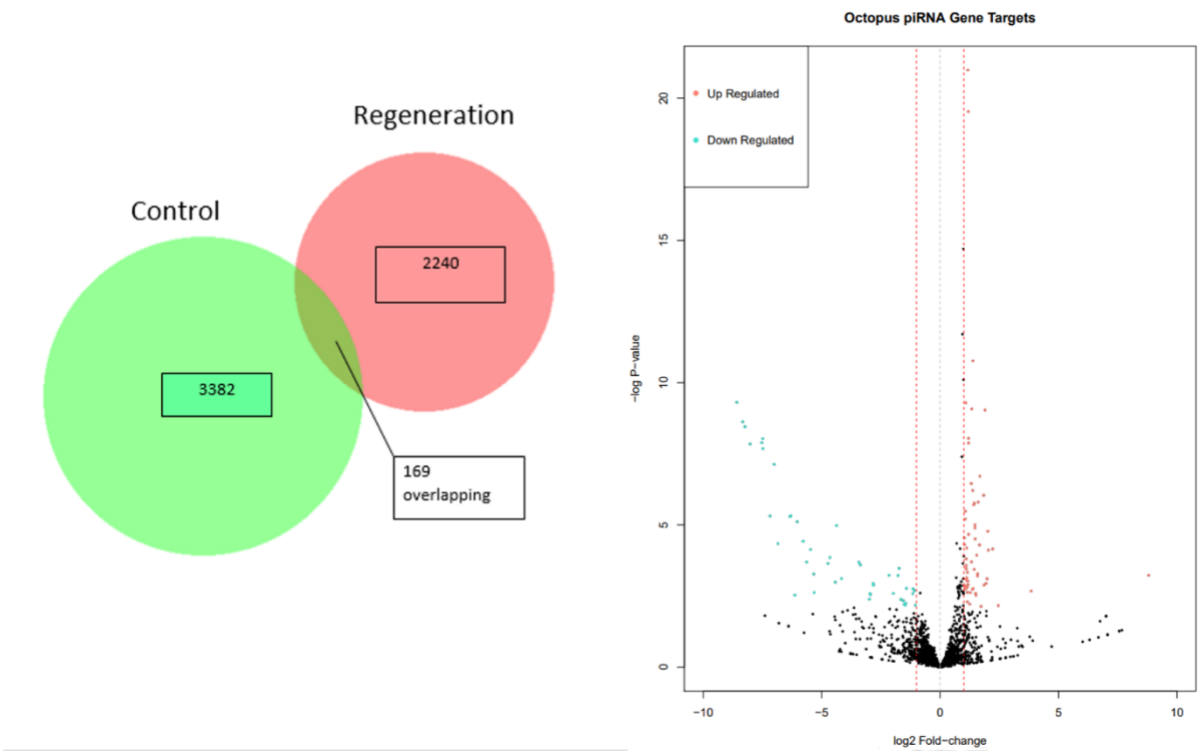

Figure 4: BioVenn Diagram of Highly Expressed Gene Loci and Differential Expression of Regeneration Only Genes. The BioVenn diagram (left) shows overlapping and unique gene loci between control (somatic) and regenerative tissue samples of O. bimaculoides*. The plot on the right shows the up-regulation and down-regulation of* O. bimaculoides *regeneration-specific genes. This was acquired using DESeq2. Significant terms (pval < 0.1 and log fold change > 1) were used to perform GO analysis (figure by author).*

Gene ontology analysis of piRNA gene targets

Gene ontology analysis was conducted on the up-regulated genes to determine the function of the genes targeted by piRNAs. The majority of targeted genes were involved in biological processes (BP), as 18 of the 30 statistically significant gene functions targeted by piRNAs were classified as BP (Fig. 5). For example, these BP gene functions

included the regulation of RNA biosynthesis, regulation of DNA-templated transcription, regulation of nucleic acid-templated translation, cell communication, and more (Fig. 5). The regulation of genetic material is crucial to regeneration to ensure the functionality of regenerated cells, while communication and signaling are essential for the determination of cell type and location. The second most heavily target gene functions were classified as molecular function (MF), as piRNAs targeted 10 MF gene functions (Fig. 5). These include regulation of numerous enzymes, such as isomerase and cholinesterase, and protein function (Fig. 5). Targeted genes were only implicated in two gene functions within the cellular component (CC) class: intrinsic component of membrane and integral component of membrane (Fig. 5).

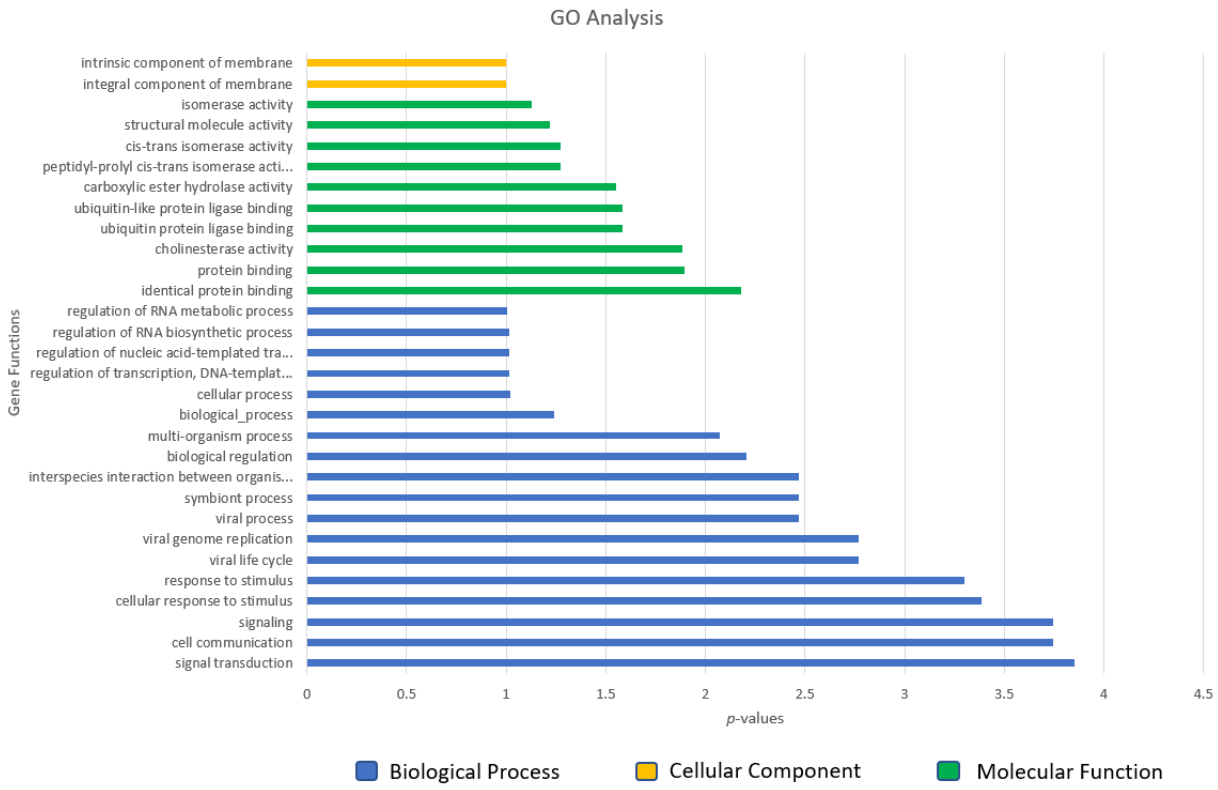

Figure 5: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) gene targeting by piRNAs in the tentacles of O. bimaculoides. *Genes that were targeted by piRNAs and their statistical significance are presented (figure by author).*

RepeatMasker analysis

A file containing regions of high expression of piRNAs that did not intersect with known annotated genes was put through RepeatMasker. This provided data on the total percentage of repeated elements, as well as the specific types of repeated elements targeted by piRNAs in the regenerative tissue from the *O. bimaculoides* tentacles (Fig. 6). Of the repeated elements, piRNAs most heavily targeted DNA transposons, long interspersed nuclear elements (LINEs), and long tandem repeats (LTRs) (Fig. 6). The

most heavily targeted DNA transposons were Tc1-IS630-Pogo and PiggyBac, and the most heavily targeted LTRs were Gypsy and DIRS1.

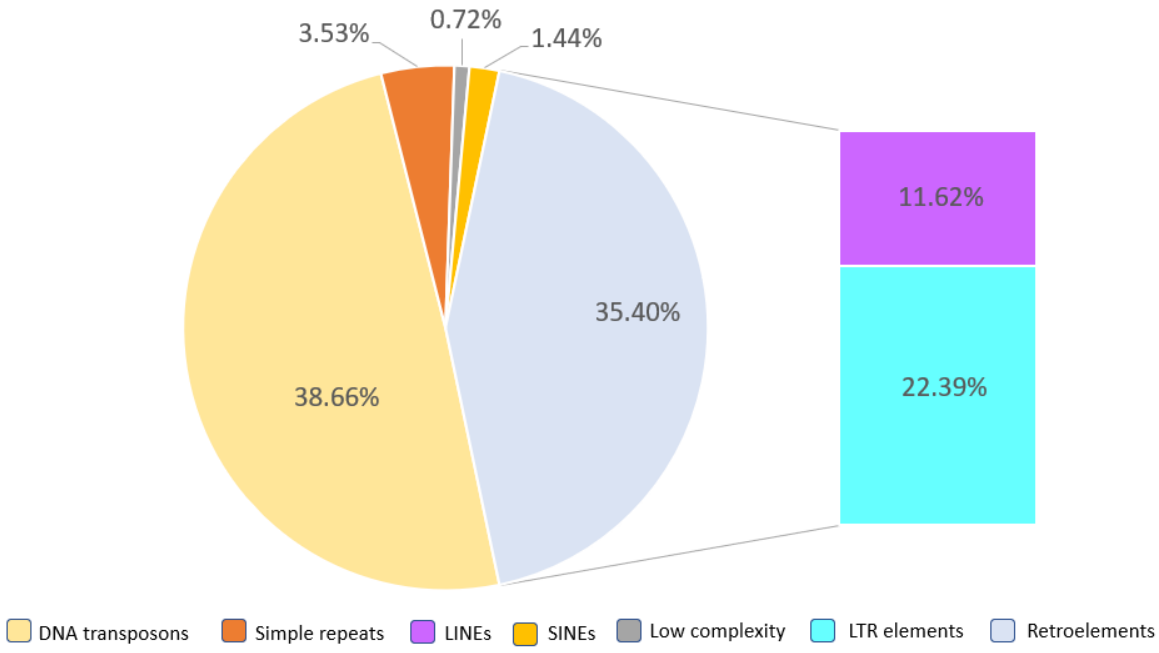

Figure 6: Transposable elements targeted by piRNAs in regenerative tissue of O. bimaculoides. *This bar of pie graph shows the percentage of piRNA targets that were not present as genes in differentially expressed loci. The retroelements portion of the pie graph is further subdivided into long tandem repeats (LTR) and long interspersed nuclear elements (LINEs) (figure by author).*

CHAPTER V: DISCUSSION

Although PIWI protein and piRNA expression has previously been briefly examined in mollusks (Jehn et al., 2018), piRNA expression has not been studied in *O. bimaculoides* specifically. In this research, differential piRNA expression was observed between the control (somatic) and regenerative tissue of *O. bimaculoides* tentacles.

Gene ontology (GO) analysis was used to further examine the functions of the gene targets of piRNAs. As mentioned previously, there are three separate possible ontologies in GO: biological process (BP), cellular component (CC), and molecular function (CC). There were 18 statistically significant BP gene functions targeted. For example, these targeted genes exhibited functions in regulation of RNA biosynthetic processes, regulation of nucleic acid-templated transcription, regulation of DNA-templated transcription, cell communication, and signal transduction, and more. These BP gene functions are directly related to the process of regeneration, as regulation of genetic material and cell signaling are both critical aspects that aid regeneration by ensuring the proper maintenance of the germline genome and informing the regenerated cells of where to locate and what type of cell to differentiate into. There were 10 statistically significant MF gene functions targeted. For example, these targeted genes exhibited functions of protein binding, identical protein binding, and the activity of numerous enzymes. The regulation of enzyme function is critical to the regenerative process due to its regulation of metabolic activity in newly formed cells. Protein binding is relevant to regeneration due to the many roles of intracellular proteins, such as their contribution to and maintenance of cell structure. There were only two statistically significant CC genes targeted: intrinsic component of membrane and integral component

of membrane. Although it was surprising that there were not more CC gene functions targeted, the intrinsic and integral components of the cell membrane are critical to regeneration, as they function in signaling, trans-membrane transport, etc., which are essential to the survival of the newly formed cells.

RepeatMasker was used to analyze the unannotated regions of the RNA sequences that did not intersect with known gene loci. This analysis showed that piRNAs were targeting repeated sequences. RepeatMasker showed that the most heavily targeted repeated sequences were DNA transposons. Long tandem repeat (LTR) retrotransposons and long interspersed nuclear elements (LINEs) were also frequently targeted. Of the LTR elements targeted by piRNAs, the Gypsy and DIRS1 groups were most heavily targeted. The most heavily targeted DNA transposons were Tc1-IS630-Pogo and PiggyBac. This targeting and silencing makes sense in regenerative tissue due to the known function of PIWI-piRNAs in genome maintenance. These transposable elements could potentially cause germline mutations in the regenerating tissue, producing defective cells and further tissue damage. Silencing of these elements ensures that new cells are functional. It is interesting that DNA transposons were targeted more than retroelements, as retroelements are by far the most common repeated elements in the genome. This suggests that piRNAs may have been preferentially targeting DNA transposons. However, the targeting of transposable elements by piRNAs was not surprising, as transposable element targeting and silencing is one of the major known functions of piRNA. This contributes to the existing knowledge of piRNA function in transposon silencing.

The data from this research suggests that piRNAs may be implicated in the impressive regenerative abilities of *O. bimaculoides*. However, further research would be required to solidify this role of piRNAs in regeneration. Such research could resemble previous studies in which mechanisms were used to suppress the function of piRNAs, followed by examination of the effects of this suppression. More research should also be conducted regarding the genes and their functions found in the GO analysis. Further examination of these genes and their function could more tightly connect piRNA expression to regeneration. It would be interesting to see this research conducted again to compare the results. The statistically significant targeting of just two CC gene functions by piRNA was surprising. Replication of this research could find more significant targeting of CC gene functions by piRNAs.

This research in *O. bimaculoides* is critical to the development of a comprehensive understanding of the process of regeneration. Animals across Kingdom Animalia maintain some level of regenerative ability (Poss, 2010). Impressive regenerative capacity is exhibited from relatively simple life forms such as *Hydra* to relatively complex creatures such as *O. bimaculoides* (Poss, 2010). Examination of PIWI proteins, piRNAs, and other mechanisms that potentially influence the process of regeneration is key to creating true regenerative medicine. The development of medicine with the ability to regenerate tissues will be a major break through in biological science, as it could be used to treat detrimental maladies such as amputated appendages, failing organs, cancer, and more.

REFERENCES

Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., ... & Rokhsar, D. S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, *524*(7564), 220-224.

Allcock, L., Taite, M. & Headlam, J. 2018. *Octopus bimaculoides*. The IUCN Red List of Threatened Species 2018: e.T163004A963213.

Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*, *318*(5851), 761–764.

Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, *409*(6818), 363-366.

Bhat, A., Ghatage, T., Bhan, S., Lahane, G. P., Dhar, A., Kumar, R., ... & Pandita, T. K. (2022). Role of transposable elements in genome stability: Implications for health and disease. *International journal of molecular sciences*, *23*(14), 7802.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089–1103.

Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S. C. R., & Lin, H. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes & Development*, *21*(18), 2300–2311.

Carmell, M. A., Xuan, Z., Zhang, M. Q., & Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & Development*, *16*(21), 2733–2742.

Cerutti, L., Mian, N., & Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends in Biochemical Sciences*, *25*(10), 481–482.

Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., & Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & development*, *12*(23), 3715-3727.

Czech, B., & Hannon, G. J. (2016). One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends in biochemical sciences*, *41*(4), 324-337.

du Plessis, L., Skunca, N., & Dessimoz, C. (2011). The what, where, how and why of gene ontology--a primer for bioinformaticians. *Briefings in Bioinformatics*, *12*(6), 723–735.

Goodwin, T. J., & Poulter, R. T. (2001). The DIRS1 group of retrotransposons. *Molecular biology and evolution*, *18*(11), 2067-2082.

Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinformatics and Biology Insights*, *9s1*, BBI.S28991.

Handberg-Thorsager, M., Fernandez, E., & Salo, E. . (2008). Stem cells and regeneration in planarians. *Frontiers in Bioscience-Landmark*, *13*(16), 6374–6394.

Harris, M., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., ... & Seigfried, T. (2004). White Ra: The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, *32*, 258-261.

Hofmeister, J. K. K. (2015). *Movement, abundance patterns, and foraging ecology of the California two spot Octopus, Octopus bimaculatus* [Doctoral dissertation, University of California at Berkeley]. ProQuest ID: Hofmeister_berkeley_0028E_15763. Merritt ID: ark:/13030/m5q28n48.

Huang, X., Tóth, K. F., & Aravin, A. A. (2017). piRNA biogenesis in *Drosophila melanogaster*. *Trends in Genetics*, *33*(11), 882-894.

Imperadore, P., & Fiorito, G. (2018). Cephalopod tissue regeneration: consolidating over a century of knowledge. *Frontiers in Physiology*, *9*, 593.

Jehn, J., Gebert, D., Pipilescu, F., Stern, S., Kiefer, J. S. T., Hewel, C., & Rosenkranz, D. (2018). PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Communications Biology*, *1*(1), 137.

Jereb, P., & Roper, C. F. (2010). *Cephalopods of the world-an annotated and illustrated catalogue of cephalopod species known to date. Vol 2. Myopsid and oegopsid squids* (No. 2). Fao.

Juliano, C., Wang, J., & Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annual Review of Genetics*, *45*(1), 447–469.

Klattenhoff, C., & Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. *Development*, *135*(1), 3–9.

Lange, M. M. (1920). On the regeneration and finer structure of the arms of the cephalopods. *Journal of Experimental Zoology*, *31*(1), 1-57.

Malik, H. S., Burke, W. D., & Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*, *16*(6), 793–805.

Pellettieri, J. (2019). Regenerative tissue remodeling in planarians – The mysteries of morphallaxis. *Seminars in Cell & Developmental Biology*, *87*, 13–21.

Percharde, M., Sultana, T., & Ramalho-Santos, M. (2020). What doesn't kill you makes you stronger: Transposons as dual players in chromatin regulation and genomic variation. *BioEssays*, *42*(4), 1900232.

Pickford, G.E.; McConnaughey, B.H. (1949). The *Octopus bimaculatus* problem: a study in sibling species. *Bulletin of the Bingham Oceanographic Collection*. 12: 1–66.

Platt, R. N., Vandewege, M. W., & Ray, D. A. (2018). Mammalian transposable elements and their impacts on genome evolution. *Chromosome Research*, *26*(1–2), 25–43.

Poss, K. D. (2010). Advances in understanding tissue regenerative capacity and mechanisms in animals. *Nature Reviews Genetics*, *11*(10), 710–722.

Poss, K. D., Keating, M. T., & Nechiporuk, A. (2003). Tales of regeneration in zebrafish. *Developmental Dynamics*, *226*(2), 202–210.

Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C., & Alvarado, A. S. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science*, *310*(5752), 1327–1330.

Ross, R. J., Weiner, M. M., & Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*, *505*(7483), 353–359.

Rossi, L., Salvetti, A., Batistoni, R., Deri, P., & Gremigni, V. (2008). Molecular and cellular basis of regeneration and tissue repair. *Cellular and Molecular Life Sciences*, *65*(1), 16–23.

Saló, E., & Baguñà, J. (2002). Regeneration in planarians and other worms: New findings, new tools, and new perspectives. *Journal of Experimental Zoology*, *292*(6), 528–539.

Shaw, T. J., Osborne, M., Ponte, G., Fiorito, G., & Andrews, P. L. R. (2016). Mechanisms of wound closure following acute arm injury in *Octopus vulgaris*. *Zoological Letters*, *2*(1), 8.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631-656.

Thomson, T., & Lin, H. (2009). The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annual Review of Cell and Developmental Biology*, *25*(1), 355–376.

Waldron, F. M., Stone, G. N., & Obbard, D. J. (2018). Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS Genetics*, *14*(7), e1007533.

Yamaguchi, S., Oe, A., Nishida, K. M., Yamashita, K., Kajiya, A., Hirano, S., ... & Nureki, O. (2020). Crystal structure of *Drosophila* piwi. *Nature communications*, *11*(1), 858.